Introduction to ILDG

Hands-on Workshop

Francesco Di Renzo - UNIPR & INFN Parma
on behalf of ILDG Board and Working Groups

June 14, 2023

Overview

1. Motivation and Objectives of ILDG
2. Beginning of ILDG
3. Basic ILDG Concepts
4. Decline and Transition to ILDG 2

Motivation and Objectives of ILDG

Gauge configurations in Lattice simulatons are valuable and expensive in terms of

- computing resources (i.e. energy, CO_2, tax payer's money)
- human effort

Objectives of ILDG

➜ make these precious raw data sharable, usable, and citable for the community

➜ ensure basic quality standards for Lattice data

➜ provide a framework for the organization and putting into practice of a solid data management (-plans)

Thus ILDG can and should be beneficial for data consumers and data providers!

Caveats

ILDG is not intended as a free solution when running out of disk space

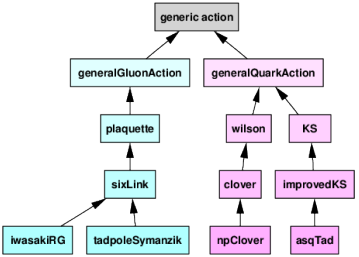
ILDG does not come for free, but also requires effort and resources on users' side

On the long run, however, this is likely to pay off, because it can

- reduce collaboration-specific efforts for island solutions
- increase quality, usability and visibility of Lattice data (and research)
- enable funding (and credit) opportunities for essential data management efforts

Beginning of ILDG

- "Data storage issues in Lattice QCD calculations" C. McNeil 2000
- ILDG proposed by R.D. Kenway in 2002
  "to manage and exchange lattice data QCD data, particularly gauge configurations"
- Concepts presented at Lattice conference in Boston 2002
- Working Groups for Metadata and Middleware starting in 2002
- ILDG Board formed in 2003
- Biannual workshops (virtual or at Lattice conferences)
- Concise and community-wide agreed metadata schema 2004
- 1st generation of federated services and infrastructure operational in 2007



Basic Concepts of ILDG

ILDG

- Federation of autonomous "Regional Grids" (RG)
- Virtual Organisation (VO)
- Community-wide agreed standards (QCDml metadata schema, data format, APIs)

ILDG operates only 2 global services:

- VO registration (VOMS)

  defines VO membership and provides VO-specific authentication attributes
- Website (temporary mirror)

  specifies standards, services and policies

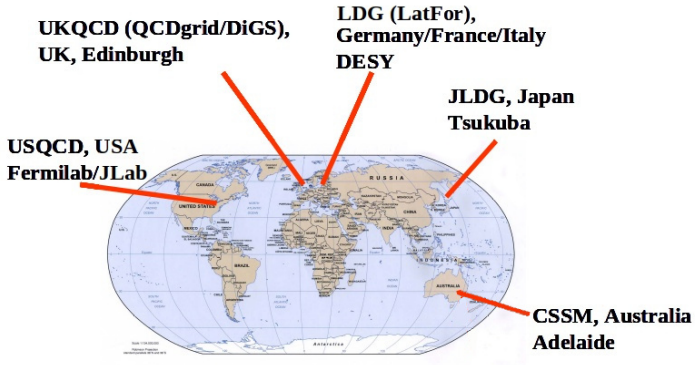
Regional Grids

CSSM, JLDG, LDG, UKQCD, USQCD

- implemented with different architectures and technologies
- organized and operated in autonomous ways with individual policies

Responsibility of each RG to provide

- Metadata Catalogue (MDC)
- File Catalogue (FC)
- Storage Elements (SE)
- website with RG-specific information



Organization of ILDG

- **Board**
  - represents ILDG towards community and service providers
  - discusses and decides policies and guidelines for membership and data sharing
  - supports regional grids in applying for resources
  - oversees working groups
- **Metadata Working Group (MDWG)**
  - agrees on concise and community-wide standards for the description of the data
  - balances scientific with technical requirements
  - provides specifications of metadata shema (QCDml) and data formats
- **Middleware Working Group (MWWG)**
  - specifies interfaces of services to ensure interoperability between all RGs
  - supports RGs by exploring technologies and sharing related expertise
  - suggests or develops prototypes of client tools e.g. for searching and download

Pioneering Efforts for ILDG

- **Board**

  R. Brower (USA), K. Jansen (Germany), R. Kenway (UK), D. Leinweber (Australia), O. Pene (France), F. Di Renzo (Italy), A. Ukawa, T. Yoshie (Japan)

- **Metadata Working Group (MDWG)**

  P. Coddington (Adelaide), T. Yoshie (Tsukuba), D. Pleiter (DESY) G. Andronico (INFN), C. Maynard (Edinburgh), C. De Tar (Utah), J. Simone (FNAL), R. Edwards, B. Joo (JLAB)

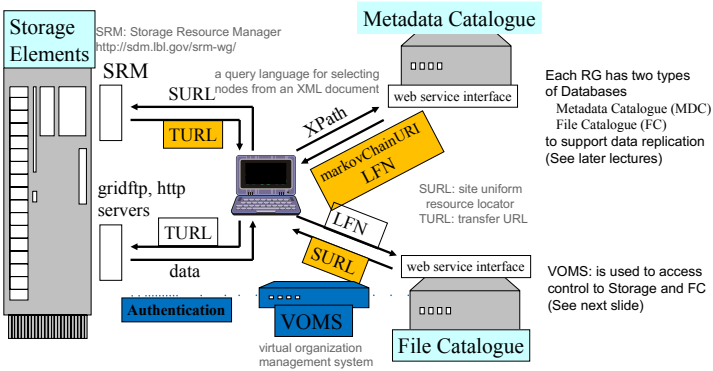
- **Middleware Working Group (MWWG)**

  P. Coddington, S. Zang (Adelaide), T. Amagasa, N. Ishii, O. Tatebe, M. Sato (Tsukuba), D. Melkumyan, D. Pleiter (DESY), G. Beckett, R. Ostrowski (Edinburgh), J. Simone (FNAL), B. Joo, C. Watson (JLAB)

List by T. Yoshie @ HackLatt2009

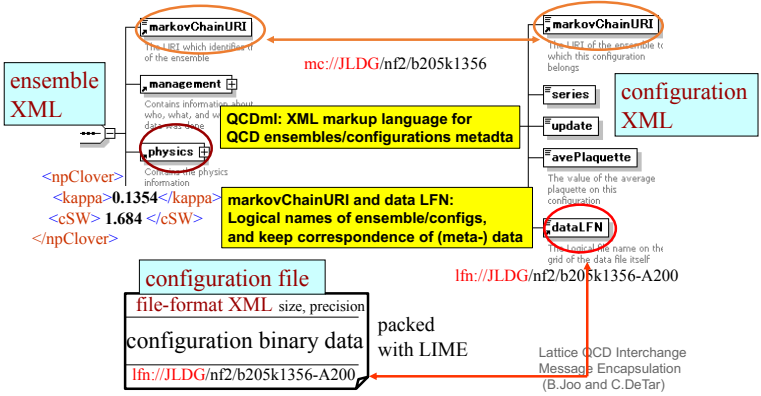
ILDG Middleware in a Nutshell

Web services of ILDG (MDC, FC, SE, VOMS) implement a distributed database system for ensemble metadata, configuration metadata, and configuration data



ILDG Metadata in a Nutshell

QCDml is a community-agreed, concisely defined, rich and flexible markup schema, which allows strict validation of metadata



Decline of ILDG

During the last 10 years:

- ✘ decreasing (rather than increasing) volume of available storage infrastructure
- ✘ degrading usability (e.g. client tools only usable by a couple of experts in those collaborations which continued to heavily use or rely on ILDG)
- ✘ no alignment with technological develoments (e.g. grid $\rightarrow$ cloud) due to loss of critical person power and expertise
- ✘ critical services going offline (often because security issues became a serious concern)
- ✘ no significant activities by Board or Working Groups after 2015

Still

- ✔ metadata and data for about 250+59 ensembles (370+30 k configs) still available in 2 RGs (but accessible only with expert knowledge)
- ✔ infrastructure still used for data sharing by individual collaborations

Towards ILDG 2

The need to resume ILDG efforts

- ✔ pile-up of data from collaborations who want to make configs publicly available (e.g. $\approx$ 3 PB only in LDG from CLS, ETMC, QCDSF)... AND YOU ARE HERE!
- ✔ Full implementation of FAIR principles. Research data managment concepts becoming incrisingly important requirement by funding agencies.

Since 2022:

- ✔ regular Working Group and Board meetings (current chair: F. Karsch)
- ✔ dedicated SW developer for re-design of Catalogue Services funded by PUNCH4NFDI (JLDG+LDG again fully operational, UK RG in preparation)
- ✔ transition from Grid to Cloud technologies
- ✔ re-activating communication within and participation by community
  - ➜ Lattice 22 (presentation and parallel session)
  - ➜ this Hands-on Workshop

Concluding Remarks

- Very diverse perspectives on (and use-cases of) ILDG
  - data consumer vs. data provider
  - restricted (within or between collaborations) vs. public sharing
- ILDG is not a commercial service
  - can only be realized as community effort
  - uni-directional propagaton of knowledge and efforts can not work
- ILDG 2 is work in progress
  - still a long way to go
  - needs replacement of components and adiabatic changes of setup
  - your feedback and participation is crucial to further develop ILDG